

NASA TECHNICAL NOTE



NASA TN D-6472

NASA TN D-6472

A STATISTICAL TECHNIQUE
FOR COMPUTER IDENTIFICATION
OF OUTLIERS IN MULTIVARIATE DATA

by

Ram Swaroop

Computing and Software, Inc.

Field Team at Flight Research Center

and

William R. Winter

Flight Research Center

Edwards, Calif. 93523

1. Report No. NASA TN D-6472	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle A STATISTICAL TECHNIQUE FOR COMPUTER IDENTIFICATION OF OUTLIERS IN MULTIVARIATE DATA		5. Report Date August 1971	
		6. Performing Organization Code	
7. Author(s) Ram Swaroop and William R. Winter		8. Performing Organization Report No. H-657	
9. Performing Organization Name and Address NASA Flight Research Center P. O. Box 273 Edwards, California 93523		10. Work Unit No. 127-49-20-00-24	
		11. Contract or Grant No.	
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, D. C. 20546		13. Type of Report and Period Covered Technical Note	
		14. Sponsoring Agency Code	
15. Supplementary Notes			
16. Abstract <p>A statistical technique and the necessary computer program for editing multivariate data are presented. The technique is particularly useful when large quantities of data are collected and the editing must be performed by automatic means. One task in the editing process is the identification of outliers, or observations which deviate markedly from the rest of the sample. A statistical technique, and the related computer program, for identifying the outliers in univariate data was presented in NASA TN D-5275. The current report is a multivariate analog which considers the statistical linear relationship between the variables in identifying the outliers. The program requires as inputs the number of variables, the data set, and the level of significance at which outliers are to be identified. It is assumed that the data are from a multivariate normal population and the sample size is at least two greater than the number of variables.</p> <p>Although the technique has been used primarily in editing biodata, the method is applicable to any multivariate data encountered in engineering and the physical sciences.</p> <p>An example is presented to illustrate the technique.</p>			
17. Key Words (Suggested by Author(s)) Outlier Multivariate outlier technique Data editing		18. Distribution Statement Unclassified - Unlimited	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 29	22. Price* \$3.00

A STATISTICAL TECHNIQUE FOR COMPUTER IDENTIFICATION OF OUTLIERS IN MULTIVARIATE DATA

Ram Swaroop
Computing and Software, Inc.
Field Team at Flight Research Center

and

William R. Winter
Flight Research Center

INTRODUCTION

The NASA Flight Research Center is engaged in an extensive biomedical research and development program. Objectives of this program include advancing the state of the art in the medical monitoring of humans in flight (ref. 1); predicting and extending the limit of man's operational capacity in the flight environment; and developing improved protection, restraint, and life support systems. As a result of this program, large quantities of biomedical information are collected in flight, necessitating dependence on the Flight Research Center's capacity for collecting, reducing, and analyzing these data by automatic means.

Experience has shown that no matter how sophisticated the monitoring, collection, and reduction systems, some editing of the biodata is required before they can be analyzed statistically. The reduced biodata may contain observations that deviate markedly from the rest of the sample. Such observations may be due to errors other than the usual random fluctuations characterizing the population to which the data belong, or may merely occur too infrequently to be considered in a particular analysis. If, upon examination, an observation falls outside a standardized region, it is usually identified as an outlier. Outliers often provide useful information. Their identification is important not only for improving the analysis but also for indicating anomalies which may require further investigation.

A statistical technique, and the related computer program, for identifying the outliers in univariate data was presented in reference 2. A method for identifying outliers in multivariate data is derived and demonstrated in this report. This method was chosen because of its simplicity and applicability in editing biodata. A program for automatic editing was written in FORTRAN IV. Inputs to this program are the number of variables, the data set, and the selected level of significance. An example is presented to illustrate the use of the method, and a scatter plot of the data is shown. The program source listing, user instructions, and a sample output are also presented.

The program computes and prints the means and standard deviations of all the variables before and after the outliers are identified and deleted. A list of the data with outliers identified by asterisks is also printed.

The authors would like to acknowledge the assistance of M. C. Nesel in writing the computer program.

SYMBOLS

A	nonsingular matrix
$F_{\alpha;p,n-p-1}$	α -level value of F-distribution with p and $(n - p - 1)$ degrees of freedom
G	normal component of acceleration as experienced by the subject, g
H/R	heart rate, beats per minute
I	identity matrix
\sum_i^k	summation starting from i through k , where i and k are integers between 1 and n , and i is less than k
$N_p(\mu, \Gamma)$	p -variate normal distribution with mean, μ , and covariance matrix, Γ
n	sample size
p	number of variables
S	$(p \times p)$ matrix of sums and cross-products of deviations of observations from \bar{X} divided by $(n - 1)$
$S.D.$	standard deviation
T^2	Hotelling's T^2 statistic
u, v	column vectors of p dimensions
X_i, X_j	i th or j th observation vector of p dimensions, where i or j ranges from 1 to n
\bar{X}	mean vector computed from n observation vectors
Z_i	i th vector obtained by orthogonal transformation of vector X_i
α	level of significance
Δ_i	positive real number corresponding to observation vector X_i , where i ranges from 1 to n
Δ_*	positive real number computed from $F_{\alpha;p, n - p - 1}$ for the data set, to compare with Δ_i

τ^2 random variable related to T^2

Superscript:

T transpose

BRIEF DESCRIPTION OF TECHNIQUE

Outliers are identified by computing, at the given level of significance, the critical value, Δ_* , for the data set and Δ_i for each observation vector, X_i . If Δ_i is larger than Δ_* , observation X_i is identified as an outlier. The quantity Δ_* is a function of total sample size, n , number of variables, p , and the F -value for the given level of significance, whereas each Δ_i is a function of the observation X_i and the estimated mean and covariance matrix from all the observations. It is assumed that all the observations constitute a random sample from a p -variate normal distribution.

DERIVATION OF TECHNIQUE

Let X_1, X_2, \dots, X_n be a random sample of size n from a p -dimensional normal distribution, $N_p(\mu, \Gamma)$. The observations will be considered as n greater than $p + 2$ column vectors in a p -dimensional vector space. Consider any $(n \times n)$ orthogonal matrix, with first two rows as shown,

$$\begin{bmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \dots & \frac{1}{\sqrt{n}} \\ \sqrt{\frac{n-1}{n}} & -\frac{1}{\sqrt{n(n-1)}} & \dots & -\frac{1}{\sqrt{n(n-1)}} \\ . & . & \dots & . \\ . & . & \dots & . \end{bmatrix}$$

representing the rotation of n -dimensional space so that the observations X_1, X_2, \dots, X_n are transformed into vectors Z_1, Z_2, \dots, Z_n , where

$$Z_1 = \sum_{i=1}^n \frac{1}{\sqrt{n}} X_i = \sqrt{n} \bar{X}$$

$$Z_2 = \sqrt{\frac{n-1}{n}} X_1 - \sum_{i=2}^n \frac{1}{\sqrt{n(n-1)}} X_i = \sqrt{\frac{n}{n-1}} (X_1 - \bar{X})$$

It may be noted that Z_1 is distributed as $N_p(\sqrt{n}\mu, \Gamma)$, Z_2, Z_3, \dots, Z_n are all distributed as $N_p(0, \Gamma)$, and all are stochastically independent of one another (ref. 3, pp. 50-52). Let S denote the estimate of the covariance matrix Γ . Then the following relation holds:

$$(n - 1)S = \sum_1^n (X_i - \bar{X})(X_i - \bar{X})^T = \sum_2^n Z_i Z_i^T$$

Define a $(p \times p)$ matrix, S_0 , such that

$$(n - 2)S_0 = \sum_3^n Z_i Z_i^T$$

Independence of Z_2 and set (Z_3, \dots, Z_n) implies that Z_2 is independent of S_0 and the

$$T^2 = Z_2^T S_0^{-1} Z_2 \quad (1)$$

statistic is distributed as Hotelling's T^2 . From the relationship between T^2 and F (ref. 3, pp. 106-107), it follows that

$$T^2 = Z_2^T S_0^{-1} Z_2$$

is distributed as

$$\frac{p(n - 2)}{(n - p - 1)} F_{p, n - p - 1}$$

Because Z_2 is not independent of S , the preceding distribution does not hold for

$$\tau^2 = Z_2^T S^{-1} Z_2 \quad (2)$$

and the distribution of τ^2 must be derived.

By the preceding definitions

$$(n - 1)S = Z_2 Z_2^T + \sum_3^n Z_i Z_i^T = Z_2 Z_2^T + (n - 2)S_0$$

or

$$(n - 2)S_o = (n - 1)S - Z_2 Z_2^T \quad (3)$$

To express the relation between T^2 and τ^2 the following lemma is used:

Lemma: Let A be a $(p \times p)$ nonsingular matrix and u, v be p -dimensional vectors. Then

$$(A - uv^T)^{-1} = A^{-1} + \frac{(A^{-1}u)(v^T A^{-1})}{1 - v^T A^{-1}u} \quad (4)$$

Proof: The proof of the lemma is presented in appendix A.

Applying the result (eq. (4)) of the lemma to equation (3),

$$S_o^{-1} = \frac{n - 2}{n - 1} \left[S^{-1} + \frac{S^{-1} Z_2 Z_2^T S^{-1}}{(n - 1) - Z_2^T S^{-1} Z_2} \right]$$

Substituting this expression for S_o^{-1} in equation (1) and applying equation (2),

$$\begin{aligned} T^2 &= Z_2^T S_o^{-1} Z_2 = \frac{n - 2}{n - 1} Z_2^T \left[S^{-1} + \frac{S^{-1} Z_2 Z_2^T S^{-1}}{(n - 1) - Z_2^T S^{-1} Z_2} \right] Z_2 \\ &= \frac{n - 2}{n - 1} \left[Z_2^T S^{-1} Z_2 + \frac{1}{(n - 1) - Z_2^T S^{-1} Z_2} (Z_2^T S^{-1} Z_2 Z_2^T S^{-1} Z_2) \right] \\ &= \frac{n - 2}{n - 1} \left[\tau^2 + \frac{\tau^4}{(n - 1) - \tau^2} \right] \\ &= \frac{(n - 2)\tau^2}{(n - 1) - \tau^2} \end{aligned} \quad (5)$$

This relation provides the distribution of τ^2 , and appropriate probability statements can be made.

Define

$$\Delta_i = (X_i - \bar{X})^T S^{-1} (X_i - \bar{X}) \text{ for } i = 1, 2, \dots, n$$

With no loss in generality, Δ_1 is used. From equation (2)

$$\tau^2 = Z_2^T S^{-1} Z_2 = \frac{n}{n-1} \Delta_1$$

and from equation (5)

$$\begin{aligned} T^2 &= \frac{(n-2) \frac{n}{n-1} \Delta_1}{(n-1) - \frac{n}{n-1} \Delta_1} \\ &= \frac{n(n-2)\Delta_1}{(n-1)^2 - n\Delta_1} \end{aligned}$$

is distributed as

$$\frac{p(n-2)}{(n-p-1)} F_{p, n-p-1}$$

From the distribution of T^2 , the statement

$$\text{Probability} \left[\frac{n(n-2)\Delta_1}{(n-1)^2 - n\Delta_1} \geq \frac{p(n-2)}{(n-p-1)} F_{\alpha; p, n-p-1} \right] = \alpha$$

provides criteria for identifying the Z_2 (or X_1) as an outlier at the assigned level of significance, α . This statement is equivalent to

$$\text{Probability} \left[\Delta_1 \geq \frac{p(n-1)^2 F_{\alpha; p, n-p-1}}{n(n-p-1) + np F_{\alpha; p, n-p-1}} \right] = \alpha$$

For significance level α , denote

$$\Delta_* = \frac{p(n-1)^2 F_{\alpha; p, n-p-1}}{n(n-p-1) + np F_{\alpha; p, n-p-1}}$$

then X_1 will be identified as an outlier at α level if

$$\Delta_1 > \Delta_*$$

The quantity X_1 (or Z_2) was chosen for convenience of the preceding derivation and the derivation holds for all X_i . Thus X_i will be identified as an outlier at α level of significance if $\Delta_i > \Delta_*$.

PROGRAM APPLICATION

Given a sample of data vectors X_1, X_2, \dots, X_n , the mean vector

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

and the estimate of the covariance matrix

$$S = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T$$

is computed. Then, for assigned α , the critical value

$$\Delta_* = \frac{p(n-1)^2 F_{\alpha; p, n-p-1}}{n(n-p-1) + np F_{\alpha; p, n-p-1}}$$

for the data set is computed. Corresponding to each observation vector, X_i ,

$$\Delta_i = (X_i - \bar{X})^T S^{-1} (X_i - \bar{X})$$

is computed. If $\Delta_i > \Delta_*$, observation vector X_i is identified as an outlier at level α .

The program (appendix B) follows this technique. The output (appendix C) of the program contains the data set, Δ_i , Δ_* , outliers marked by asterisks (*), the number of outliers identified, and the level of significance. The output also shows the means and standard deviations of the variables before and after the deletion of outliers. The required input parameters are: (1) format of the data to be read, (2) number of variables, (3) significance level, α , and (4) the data set, formatted as specified. Program options allow the user to select either a 5 percent or a 1 percent level of significance and to print the names of the variables, if desired. This program is designed so that it can be used as a subroutine in other than biodata applications, in engineering and the physical sciences, for example.

The program is particularly useful when large quantities of data are collected and the editing must be performed by automatic means.

EXAMPLE

Heart rate, H/R , and normal component of acceleration, G , data from a 66-minute flight by a student pilot at the Aerospace Research Pilot School, Edwards Air Force Base, Calif., are used to demonstrate the described technique of computer editing of biodata. These data were chosen because centrifuge studies (ref. 4) have shown that H/R and G are linearly related. The program was used to identify the outliers at a 1 percent level of significance considering H/R and G separately as univariate data and together as bivariate data. The computer output for these cases is shown in appendix C.

The results of the two univariate analyses and the one bivariate analysis of the same data are presented in figure 1. The point labeled H is identified as an outlier on the basis of H/R analysis alone; the point labeled G is identified as an outlier on the basis of G alone; and points labeled B are identified as outliers on the basis of bivariate analysis of H/R and G .

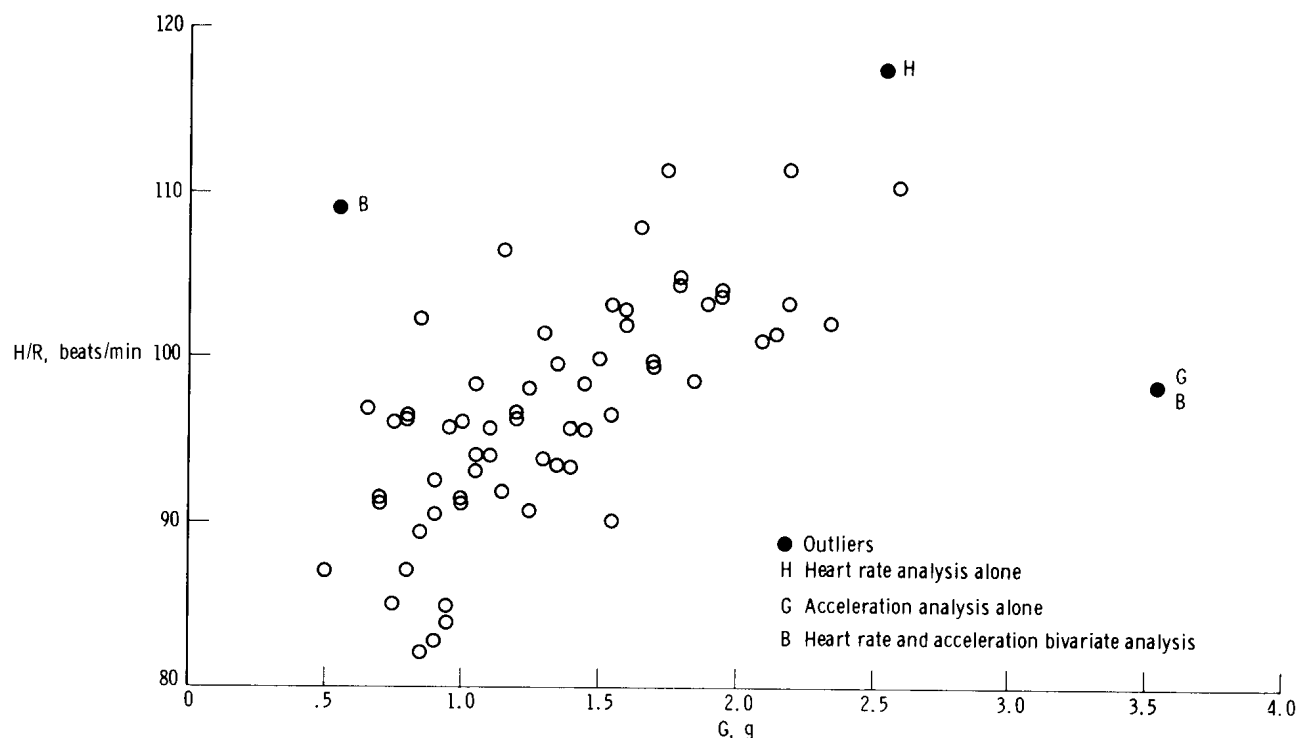


Figure 1. Minute heart rate and acceleration data for a 66-minute flight of a student pilot from the Aerospace Research Pilot School showing outliers identified by the automatic multivariate outlier technique.

Bivariate analysis is based on the fact that high H/R is associated with high G , whereas univariate analysis cannot take this information into account. For this reason the point labeled H was not identified by the bivariate analysis, but was identified as an outlier on the basis of univariate H/R analysis. Also, both points labeled B appear not to follow a statistical linear relationship and are identified by the bivariate

analysis; however, only one of these points was identified by one of the univariate analyses (G alone). This example thus focuses on the fact that the multivariate technique, which utilizes the statistical linear relationships between the variables, is preferable in identifying outliers in multivariate data.

CONCLUDING REMARKS

A statistical technique to identify outliers, or observations which deviate markedly from the rest of the sample, in multivariate data at a given level of significance was derived. The use of the technique was illustrated by a biodata example. The example also demonstrated that the results obtained when each variable was considered separately could be different from the results obtained when the variables were considered jointly. The latter technique takes into account the statistical linear relationship between the variables and is the preferred method.

Although this method of detecting and identifying outliers is being used for biodata editing at the NASA Flight Research Center, it is also applicable to multivariate data encountered in other disciplines, such as engineering and the physical sciences. This technique is particularly useful when large quantities of data are collected and the editing must be performed by automatic means.

The program can be used as a subroutine in multivariate analyses.

Flight Research Center,

National Aeronautics and Space Administration,

Edwards, Calif., May 5, 1971.

APPENDIX A

PROOF OF THE LEMMA

Lemma: If A is a $(p \times p)$ nonsingular matrix, and u, v are p -dimensional vectors, then

$$(A - uv^T)^{-1} = A^{-1} + \frac{(A^{-1}u)(v^T A^{-1})}{1 - v^T A^{-1}u}$$

Proof: The result is obtained by showing that

$$(A - uv^T) \left[A^{-1} + \frac{(A^{-1}u)(v^T A^{-1})}{1 - v^T A^{-1}u} \right] = I$$

Simplification of the left-hand expression gives

$$AA^{-1} + \frac{AA^{-1}u(v^T A^{-1})}{1 - v^T A^{-1}u} - uv^T A^{-1} - \frac{uv^T A^{-1}uv^T A^{-1}}{1 - v^T A^{-1}u}$$

or

$$I + \frac{1}{1 - v^T A^{-1}u} \left[uv^T A^{-1} - uv^T A^{-1} + (v^T A^{-1}u)(uv^T A^{-1}) - uv^T A^{-1}uv^T A^{-1} \right]$$

Because $v^T A^{-1}u$ is a scalar, the expression becomes

$$I + \frac{1}{1 - v^T A^{-1}u} \left[(v^T A^{-1}u)(uv^T A^{-1}) - (v^T A^{-1}u)(uv^T A^{-1}) \right]$$

or

$$I$$

which is the same as the right-hand side.

APPENDIX B

PROGRAM SOURCE LISTING

```

C
C      MTVOUT
C
C      PURPOSE
C          IDENTIFY OUTLIERS IN MULTIVARIATE DATA
C
C      METHOD
C          LET VECTORS X(1) THRU X(N) BE OBSERVATIONS FROM A NP-VARIATE
C          NORMAL DISTRIBUTION
C          THEN
C              XBAR AND COVARIANCE MATRIX S ARE OBTAINED BY EQUATIONS
C              
$$XBAR = (1/N) * \sum X(J)$$

C              
$$S * (N-1) = \sum (X(J) - XBAR) * (X(J) - XBAR) TRANSPOSE$$

C
C          WHERE
C              N = SAMPLE SIZE          (N.LE.500)
C              NP = NUMBER OF VARIABLES (NP.LE.10)
C              ALPH = SIGNIFICANCE LEVEL
C              FALPH = F-VALUE FOR ALPHA AT NP AND N-NP-1
C                   DEGREES OF FREEDOM
C              X(J,I) = THE 'I' TH ELEMENT OF THE 'J' TH VECTOR,
C                   WHERE I=1,2,...,NP AND J=1,2,...,N
C
C      CALCULATE
C          DELSTAR = ((N-1)**2*NP*FALPH) / (N*((N-NP-1)+NP*FALPH))
C
C      CALCULATE FOR EACH OBSERVATION VECTOR X(J)
C          RR = (X(J)-XBAR) TRANSPOSE * (S) INVERSE * (X(J)-XBAR)
C
C      IF RR > DELSTAR, THEN X(J) IS IDENTIFIED AS AN OUTLIER
C
C      REFERENCE
C
C          1. AN INTRODUCTION TO MULTIVARIATE STATISTICAL ANALYSIS,
C             ANDERSON, 1965
C          2. A SIMPLE TECHNIQUE FOR AUTOMATIC COMPUTER EDITING
C             OF BIODATA, NASA TN D-5275
C          3. SYS/360 SCIENTIFIC SUBROUTINE PACKAGE (360A-CM-03X)
C             PROGRAMMERS MANUAL, IBM INC., 1968
C
C      SUBROUTINES
C
C          F TABLE
C          MPRD          (IBM SSP)
C          LOC           "  "
C          DSINV         "  "
C          DMFSD         "  "
C

```

APPENDIX B

```

C      INPUT
C
C      CARD 1  FORMAT OF X-ARRAY CARDS TO BE READ IN          (20A4)
C      CARD 2
C          COL 1- 2      NP      (NUMBER OF VARIABLES)  NP.LE.10      (12)
C          COL 3         BLANK
C          COL 4- 6      ALPH      (SIGNIFICANCE LEVEL) .05 OR .01 (F3.2)
C          COL 7- 9      BLANK
C          COL 10        VARIABLE NAME CARD INDICATOR
C                      1 - NAME CARD FOLLOWS
C                      BLANK - NO NAME CARD
C      CARD 3  (OPTIONAL)
C          TEN FIELDS OF EIGHT CHARACTERS EACH (10A8), WHICH MAY
C          BE USED TO ASSIGN MEANINGFUL NAMES TO THE NP VARIABLES.
C          IF COL 10 OF THE PREVIOUS CARD IS PUNCHED, NAMES MUST
C          BE ASSIGNED FOR ALL NP VARIABLES.  DEFAULT NAMES ARE
C          'X1', 'X2', ... , 'X(NP)'.
C      CARDS 4-      DATA FORMATTED AS PRESCRIBED IN CARD 1
C
C      MULTIPLE RUNS ARE PERMITTED, AS LONG AS EACH DATA DECK IS
C      PRECEDED BY APPROPRIATE CONTROL CARDS (CARDS 1, 2 AND
C      3 ABOVE), AND IS FOLLOWED BY A CARD WITH **** PUNCHED
C      IN COLUMNS 1 THRU 4.
C
C      OUTPUT
C      1 LIST OF VECTORS WITH OUTLIERS IDENTIFIED BY ASTERISKS
C      2 MEAN (ORIGINAL DATA)
C      3 STANDARD DEVIATION (ORIGINAL DATA)
C      4 MEAN (OUTLIERS DELETED)
C      5 STANDARD DEVIATION (OUTLIERS DELETED)
C
C      REAL*8 SUMX(10), XBAR(10), SD(10), XBAR1(10), SD1(10)
C      REAL*8 A(55), B(10), S(10,10)
C      REAL*8 RR, DEL STR, R(10), FMT(10), BLNK/'      '/, VAR(10)
C      REAL*8 DEF(10)/'      X1  ','      X2  ','      X3  ','      X4  ','      X5
1  ','      X6  ','      X7  ','      X8  ','      X9  ','      X10 '/'
C      REAL      X(501,10), Y(501,10)
C      INTEGER STAR/'****'/, BLANK/'      '/, FLAG, FLGTOT, DSWI
C      INTEGER PREFMT(2)/'(A4,', 'T1, '/, OPT(2)/' YES', ' NC'/
C      INTEGER*2 IFLAG(500)
C      LOGICAL*1 FORMAT(89), RPAREN/'(')/
C      EQUIVALENCE (PREFMT,FORMAT(1)),(FMT,FORMAT(9)),(RPAREN,FORMAT(89))
1  DO 2  I=1,10
C      VAR(I) = DEF(I)
2  FMT(I) = BLNK
C      DO 5  I=1,500
5  IFLAG(I) = 0
C      FLGTOT = 0
C      LINES = 1
C      NSWI = 0
C      DSWI = 0
C      OH = 0.01

```

APPENDIX B

```

      IO = 1
      IZ = 0
      MM = 1
      N = 1
      READ(1,1000,END=999) FMT
      READ(1,1010) NP, ALPH, ISWI
C      TEST IF NAMES HAVE BEEN ASSIGNED TO THE VARIABLES
      IF(ISWI.GT.0) GO TO 10
      MM = 2
      GO TO 15
10  READ(1,1050) (VAR(I), I=1,NP)
C      READ IN THE DATA AS X-ARRAY
15  READ(1,FORMAT,END=995) IXX, (X(N,I), I=1,NP)
16  IF(IXX.EQ.STAR) GO TO 19
      N = N + 1
      IF(N.GT.501) GO TO 990
      GO TO 15
19  N = N - 1
C      SELECT APPROPRIATE F-VALUE
      CALL FTABLE (N,NP,ALPH,FALPH)
C      WRITE THE INPUT CONTROL INFORMATION
      WRITE(3,5002) NP, ALPH, FALPH, OPT(MM)
      DELSTR = ((N-1)**2 * NP * FALPH) / (N * ((N-NP-1) + NP * FALPH))
20  DO 86 I=1,10
      SUMX(I)=0.D0
      DO 86 J=1,10
86  S(I,J)=0.D0
      DO 30 J=1,N
C      TEST FOR FLAGGED VECTORS IDENTIFIED AS OUTLIERS
      IF(IFLAG(J).GT.0) GO TO 30
      DO 25 I=1,NP
25  SUMX(I) = SUMX(I) + X(J,I)
30  CONTINUE
C      FIND MEAN OF EACH VARIABLE
      DO 40 I=1,NP
      XBAR(I) = SUMX(I) / (N-FLGTOT)
      DO 40 J=1,N
      Y(J,I) = X(J,I) - XBAR(I)
40  IF(IFLAG(J).GT.0) Y(J,I) = 0.D0
      JJ = 0
      DO 70 I=1,NP
      DO 70 K=1,I
      DO 60 J=1,N
60  S(I,K) = Y(J,I) * Y(J,K) + S(I,K)
C      FIND STANDARD DEVIATION OF EACH VARIABLE
      SD(I) = DSQRT(S(I,I) / (N-1-FLGTOT))
      JJ = JJ + 1
70  A(JJ) = S(I,K)
C      IF COMPUTATIONS ARE COMPLETE, BRANCH TO
C      PRINT TABLE OF MEANS AND S.D.'S
      IF(DSWI.GT.0) GO TO 200
      CALL DSINV(A,NP,OH,IER)

```

APPENDIX B

```

      IF( IER ) 991,80,991
C      WRITE THE LIST HEADING
80  WRITE(3,1015) (VAR(I), I=1,NP)
      DO 100 J=1,N
      DO 90 K=1,NP
90  B(K) = Y(J,K)
      CALL MPRD (B,A,R,IO,NP,IZ,IC,NP)
      CALL MPRD (R,B,RR,IO,NP,IZ,IZ,IO)
      RR = RR*(N-1)
      FLAG = BLANK
C      FLAG THIS VECTOR WITH AN ASTERISK IF
C      IT IS IDENTIFIED AS AN OUTLIER
      IF(RR.GT.DELSTR) FLAG = STAR
      IF(FLAG.NE.STAR) GO TO 95
      IFLAG(J) = 1
      FLGTOT = FLGTOT + 1
C      WRITE THE DATA VECTOR AND IF IDENTIFIED AS
C      AN OUTLIER, LABEL WITH AN ASTERISK
95  WRITE(3,1020) J, RR, FLAG, (X(J,K), K=1,NP)
      LINES = LINES + 1
      IF(LINES.LE.55) GO TO 100
      WRITE(3,1015) (VAR(I), I=1,NP)
      LINES = 1
100  CONTINUE
      WRITE(3,1025) ALPH
      WRITE(3,1030) N
      WRITE(3,1035) FLGTOT
      WRITE(3,1040) DELSTR
C      SAVE MEANS AND S.D.'S, THEN LCCP BACK AND
C      COMPUTE NEW MEANS AND S.D.'S AFTER DELETING OUTLIERS
      DO 110 I=1,NP
      XBAR1(I) = XBAR(I)
110  SD1(I) = SD(I)
      DSWI = 1
      GO TO 20
C      WRITE TABLE OF MEANS AND S.D.'S BEFORE AND
C      AFTER DELETION OF OUTLIERS
200  WRITE(3,2000)
      WRITE(3,2005)
      WRITE(3,2010)
      DO 210 I=1,NP
210  WRITE(3,2015) VAR(I), XBAR1(I), SD1(I), XBAR(I), SD(I)
      IF(NSWI.EQ.1) GO TO 999
      GO TO 1
C
990  WRITE (3,5000)
      GO TO 9999
991  WRITE (3,5001)
      GO TO 9999
995  NSWI = 1
      GO TO 19
999  WRITE (3,5009)

```

APPENDIX B

```

1000 FORMAT (20A4)
1010 FORMAT (I2,1X,F3.2,3X,I1)
1015 FORMAT (1H1,'LIST OF VECTORS WITH OUTLIERS IDENTIFIED BY ASTERISKS
1//1H0,' J',7X,'DELTA',7X,1C(2X,A8)/)
1020 FORMAT (1H ,I3,F12.4,2X,A1,4X,10F10.2)
1025 FORMAT (///1H0,12X,'* OUTLIER IDENTIFIED AT ',F4.2,' SIGNIFICANCE
LEVEL')
1030 FORMAT (1H ,12X,'SAMPLE SIZE IS ',I3)
1035 FORMAT (1H ,12X,'NO OF OUTLIERS IS ',I3)
1040 FORMAT (1H ,12X,'DELSTAR = ',F10.4)
1050 FORMAT (10A8)
2000 FORMAT (1H1,20X,'MEAN AND STANDARD DEVIATION OF THE VARIABLES')
2005 FORMAT (1H0,14X,'DATA BEFORE IDENTIFICATION',9X,'DATA AFTER DELETI
ON'/1H ,22X,'OF OUTLIERS',21X,'OF OUTLIERS')
2010 FORMAT(1H0,'VARIABLES',8X,'MEAN',13X,'S.D.',11X,'MEAN',14X,'S.D.')
2015 FORMAT (1H0,A8,F15.4,F17.4,F15.4,F18.4)
5000 FORMAT (1H1,'SAMPLE SIZE EXCEEDS 500 - PROGRAM TERMINATED')
5001 FORMAT (1H1,'ERROR IN THE MATRIX INVERSION PROCESS - PROGRAM TERMI
NATED')
5002 FORMAT (1H1,'**INPUT CONTROL INFORMATION'/1H0,'**NUMBER OF VARIAB
LES IS',I4/1H , '**SIGNIFICANCE LEVEL IS',F5.2/1H , '**F-VALUE IS',
2F16.4//1H , '**VARIABLE NAME CARD: ',A4)
5009 FORMAT (1H1,'END OF JOB')
9999 STOP
      END

```

APPENDIX B

```

SUBROUTINE FTABLE (N,M,SL,F)
C*****
C
C      FTABLE
C
C      SUBROUTINE FTABLE SELECTS THE PROPER VALUE OF F AT M AND N-M-1
C      DEGREES OF FREEDOM FOR SIGNIFICANCE LEVELS OF 5% OR 1%.
C
C      CALLING PARAMETERS
C
C          N = SAMPLE SIZE                      INTEGER
C          M = NUMBER OF VARIABLES              M.LE.10    INTEGER
C          SL = SIGNIFICANCE LEVEL              .05 OR .01  REAL
C          F = SELECTED F-VALUE                  REAL
C
C*****
C      DIMENSION TABLE(2,10,37)
C      INTEGER XDF(7)/30,40,60,120,200,400,1000/, DF
C      REAL SLT(2)/.05,.01/
C      REAL  TABL1(200)/161.4,4052.,199.5,4999.5,215.7,5403.,224.6,5625.,
1230.2,5764.,234.0,5859.,236.8,5928.,238.9,5982.,240.5,6022.,241.9,
26056.,18.51,98.50,19.00,99.00,19.16,99.17,19.25,99.25,19.30,99.30,
319.33,59.33,19.35,99.36,19.37,99.37,19.38,99.39,19.40,99.40,10.13,
434.12, 9.55,30.82, 9.28,29.46, 9.12,28.71, 9.01,28.24, 8.94,27.91,
5 8.89,27.67, 8.85,27.49, 8.81,27.35, 8.79,27.23, 7.71,21.20, 6.94,
618.00, 6.59,16.69, 6.39,15.98, 6.26,15.52, 6.16,15.21, 6.09,14.98,
7 6.04,14.80, 6.00,14.66, 5.96,14.55, 6.61,16.28, 5.79,13.27, 5.41,
812.06, 5.19,11.39, 5.05,10.97, 4.95,10.67, 4.88,10.46, 4.82,10.29,
9 4.77,10.16, 4.74,10.05, 5.99,13.75, 5.14,10.92, 4.76, 9.78, 4.53,
1 9.15, 4.39, 8.75, 4.28, 8.47, 4.21, 8.26, 4.15, 8.10, 4.10, 7.98,
2 4.06, 7.87, 5.59,12.25, 4.74, 9.55, 4.35, 8.45, 4.12, 7.85, 3.97,
3 7.46, 3.87, 7.19, 3.79, 6.99, 3.73, 6.84, 3.68, 5.72, 3.64, 6.62,
4 5.32,11.26, 4.46, 8.65, 4.07, 7.59, 3.84, 7.01, 3.69, 6.63, 3.58,
5 6.37, 3.50, 6.18, 3.44, 6.03, 3.39, 5.91, 3.35, 5.81, 5.12,10.56,
6 4.26, 8.02, 3.86, 6.99, 3.63, 6.42, 3.48, 6.05, 3.37, 5.80, 3.29,
7 5.61, 3.23, 5.47, 3.18, 5.35, 3.14, 5.26, 4.96,10.04, 4.10, 7.56,
8 3.71, 6.55, 3.48, 5.99, 3.33, 5.64, 3.22, 5.39, 3.14, 5.20, 3.07,
9 5.06, 3.02, 4.94, 2.98, 4.85/
C      REAL  TABL2(200)/ 4.84,9.65,3.98,7.21,3.59,6.22,3.36,5.67,3.20,
15.32,3.09,5.07,3.01,4.89,2.95,4.74,2.90,4.63,2.85,4.54,4.75,9.33,
23.89,6.93,3.49,5.95,3.26,5.41,3.11,5.06,3.00,4.82,2.91,4.64,2.85,
34.50,2.80,4.39,2.75,4.30,4.67,9.07,3.81,6.70,3.41,5.74,3.18,5.21,
43.03,4.86,2.92,4.62,2.83,4.44,2.77,4.30,2.71,4.19,2.67,4.10,4.60,
58.86,3.74,6.51,3.34,5.56,3.11,5.04,2.96,4.69,2.85,4.46,2.76,4.28,
62.70,4.14,2.65,4.03,2.60,3.94,4.54,8.68,3.68,6.36,3.29,5.42,3.06,
74.89,2.90,4.56,2.75,4.32,2.71,4.14,2.64,4.00,2.59,3.89,2.54,3.80,
84.49,8.53,3.63,6.23,3.24,5.29,3.01,4.77,2.85,4.44,2.74,4.20,2.66,
94.03,2.59,3.89,2.54,3.78,2.49,3.69,4.45,8.40,3.59,6.11,3.20,5.18,
12.96,4.67,2.81,4.34,2.70,4.10,2.61,3.93,2.55,3.79,2.49,3.68,2.45,
23.59,4.41,8.29,3.55,6.01,3.16,5.09,2.93,4.58,2.77,4.25,2.66,4.01,
32.58,3.84,2.51,3.71,2.46,3.60,2.41,3.51,4.38,8.18,3.52,5.93,3.13,
45.01,2.90,4.50,2.74,4.17,2.63,3.94,2.54,3.77,2.48,3.63,2.42,3.52,

```

APPENDIX B

```

52.38,3.43,4.35,8.10,3.49,5.85,3.10,4.94,2.87,4.43,2.71,4.10,2.60, FTAB0520
63.87,2.51,3.70,2.45,3.56,2.39,3.46,2.35,3.37/ FTAB0530
REAL TABL3(200)/ 4.32,8.02,3.47,5.78,3.07,4.87,2.84,4.37,2.68, FTAB0540
14.04,2.57,3.81,2.49,3.64,2.42,3.51,2.37,3.40,2.32,3.31,4.30,7.95, FTAB0550
23.44,5.72,3.05,4.82,2.82,4.31,2.66,3.99,2.55,3.76,2.46,3.59,2.40, FTAB0560
33.45,2.34,3.35,2.30,3.26,4.28,7.88,3.42,5.66,3.03,4.76,2.80,4.26, FTAB0570
42.64,3.94,2.53,3.71,2.44,3.54,2.37,3.41,2.32,3.30,2.27,3.21,4.26, FTAB0580
57.82,3.40,5.61,3.01,4.72,2.78,4.22,2.62,3.90,2.51,3.67,2.42,3.50, FTAB0590
62.36,3.36,2.30,3.26,2.25,3.17,4.24,7.77,3.39,5.57,2.99,4.68,2.76, FTAB0600
74.18,2.60,3.85,2.49,3.63,2.40,3.46,2.34,3.32,2.28,3.22,2.24,3.13, FTAB0610
84.23,7.72,3.37,5.53,2.98,4.64,2.74,4.14,2.59,3.82,2.47,3.59,2.39, FTAB0620
93.42,2.32,3.29,2.27,3.18,2.22,3.09,4.21,7.68,3.35,5.49,2.96,4.60, FTAB0630
12.73,4.11,2.57,3.78,2.46,3.56,2.37,3.39,2.31,3.26,2.25,3.15,2.20, FTAB0640
23.06,4.20,7.64,3.34,5.45,2.95,4.57,2.71,4.07,2.56,3.75,2.45,3.53, FTAB0650
32.36,3.36,2.29,3.23,2.24,3.12,2.19,3.03,4.18,7.60,3.33,5.42,2.93, FTAB0660
44.54,2.70,4.04,2.55,3.73,2.43,3.50,2.35,3.33,2.28,3.20,2.22,3.09, FTAB0670
52.18,3.00,4.17,7.56,3.32,5.39,2.92,4.51,2.69,4.02,2.53,3.70,2.42, FTAB0680
63.47,2.33,3.30,2.27,3.17,2.21,3.07,2.16,2.98/ FTAB0690
REAL TABL4(140)/ 4.08,7.31,3.23,5.18,2.84,4.31,2.61,3.83,2.45, FTAB0700
13.51,2.34,3.29,2.25,3.12,2.18,2.99,2.12,2.89,2.08,2.80,4.00,7.08, FTAB0710
23.15,4.98,2.76,4.13,2.53,3.65,2.37,3.34,2.25,3.12,2.17,2.95,2.10, FTAB0720
32.82,2.04,2.72,1.99,2.63,3.92,6.85,3.07,4.79,2.68,3.95,2.45,3.48, FTAB0730
42.29,3.17,2.17,2.96,2.09,2.75,2.02,2.66,1.96,2.56,1.91,2.47,3.89, FTAB0740
56.76,3.04,4.71,2.65,3.88,2.41,3.41,2.26,3.11,2.14,2.90,2.05,2.73, FTAB0750
61.98,2.60,1.92,2.50,1.87,2.41,3.86,6.70,3.02,4.66,2.62,3.83,2.39, FTAB0760
73.36,2.23,3.06,2.12,2.85,2.03,2.69,1.96,2.55,1.90,2.46,1.85,2.37, FTAB0770
83.85,6.66,3.00,4.62,2.61,3.80,2.38,3.34,2.22,3.04,2.10,2.82,2.02, FTAB0780
92.66,1.95,2.53,1.89,2.43,1.84,2.34,3.84,6.63,3.00,4.61,2.60,3.78, FTAB0790
12.37,3.32,2.21,3.02,2.10,2.80,2.01,2.64,1.94,2.51,1.88,2.41,1.83, FTAB0800
22.32/ FTAB0810
EQUIVALENCE (TABLE(1,1,1),TABL1(1)), (TABLE(1,1,11),TABL2(1)) FTAB0820
EQUIVALENCE (TABLE(1,1,21),TABL3(1)), (TABLE(1,1,31),TABL4(1)) FTAB0830
IF(M.GT.10) GO TO 590 FTAB0840
DF = N-M-1 FTAB0850
IF(SL.EQ.SLT(1)) GO TO 10 FTAB0860
IF(SL.EQ.SLT(2)) GO TO 15 FTAB0870
GO TO 591 FTAB0880
10 L = 1 FTAB0890
GO TO 16 FTAB0900
15 L = 2 FTAB0910
16 IF(DF.LE.30) GO TO 40 FTAB0920
DO 20 I=2,7 FTAB0930
IF(DF-XDF(I)) 50,30,20 FTAB0940
20 CONTINUE FTAB0950
F = TABLE(L,M,36) + ((1./N)/.001) * (TABLE(L,M,36)-TABLE(L,M,37)) FTAB0960
RETURN FTAB0970
30 DF = I+29 FTAB0980
40 F = TABLE(L,M,DF) FTAB0990
RETURN FTAB1000
50 F = TABLE(L,M,I+29) + ((1./DF - 1./XDF(I)) / (1./XDF(I-1) - 1./XDF(I))) * (TABLE(L,M,I+28) - TABLE(L,M,I+29)) FTAB1010
RETURN FTAB1020
590 WRITE(3,990) FTAB1030
F = 2.32 FTAB1040
RETURN FTAB1050
591 WRITE(3,991) SL FTAB1060
SL = 0.01 FTAB1070
GO TO 15 FTAB1080
990 FORMAT (1H1,'**NUMBER OF VARIABLES > 10. F HAS BEEN SET TO A DUMM FTAB1100
1Y VALUE OF 2.32') FTAB1110
991 FORMAT (1H1,'**SIGNIFICANCE LEVEL ',F6.3,' IS NOT ACCEPTABLE. LEV FTAB1120
1EL IS SET TO .01 .') FTAB1130
END FTAB1140

```

APPENDIX B

```

C
C
C .....MPRD 001
C .....MPRD 002
C .....MPRD 003
C SUBROUTINE MPRD .....MPRD 004
C .....MPRD 005
C PURPOSE .....MPRD 006
C MULTIPLY TWO MATRICES TO FORM A RESULTANT MATRIX .....MPRD 007
C .....MPRD 008
C USAGE .....MPRD 009
C CALL MPRD%A,B,R,N,M,MSA,MSB,L .....MPRD 010
C .....MPRD 011
C DESCRIPTION OF PARAMETERS .....MPRD 012
C A - NAME OF FIRST INPLT MATRIX .....MPRD 013
C B - NAME OF SECOND INPLT MATRIX .....MPRD 014
C R - NAME OF OUTPUT MATRIX .....MPRD 015
C N - NUMBER OF ROWS IN A AND R .....MPRD 016
C M - NUMBER OF COLUMNS IN A AND ROWS IN B .....MPRD 017
C MSA - ONE DIGIT NUMBER FOR STORAGE MODE OF MATRIX A .....MPRD 018
C 0 - GENERAL .....MPRD 019
C 1 - SYMMETRIC .....MPRD 020
C 2 - DIAGONAL .....MPRD 021
C MSB - SAME AS MSA EXCEPT FOR MATRIX B .....MPRD 022
C L - NUMBER OF COLUMNS IN B AND R .....MPRD 023
C .....MPRD 024
C REMARKS .....MPRD 025
C MATRIX R CANNOT BE IN THE SAME LOCATION AS MATRICES A OR B .....MPRD 026
C NUMBER OF COLUMNS OF MATRIX A MUST BE EQUAL TO NUMBER OF ROWS .....MPRD 027
C OF MATRIX B .....MPRD 028
C .....MPRD 029
C SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED .....MPRD 030
C LOC .....MPRD 031
C .....MPRD 032
C METHOD .....MPRD 033
C THE M BY L MATRIX B IS PREMULTIPLIED BY THE N BY M MATRIX A .....MPRD 034
C AND THE RESULT IS STORED IN THE N BY L MATRIX R. THIS IS A .....MPRD 035
C ROW INTO COLUMN PRODUCT. .....MPRD 036
C THE FOLLOWING TABLE SHOWS THE STORAGE MODE OF THE OUTPUT .....MPRD 037
C MATRIX FOR ALL COMBINATIONS OF INPUT MATRICES .....MPRD 038
C
C      A          B          R          .....MPRD 039
C      GENERAL    GENERAL    GENERAL    .....MPRD 040
C      GENERAL    SYMMETRIC    GENERAL    .....MPRD 041
C      GENERAL    DIAGONAL     GENERAL    .....MPRD 042
C      SYMMETRIC   GENERAL      GENERAL    .....MPRD 043
C      SYMMETRIC   SYMMETRIC    GENERAL    .....MPRD 044
C      SYMMETRIC   DIAGONAL     GENERAL    .....MPRD 045
C      DIAGONAL    GENERAL      GENERAL    .....MPRD 046
C      DIAGONAL    SYMMETRIC    GENERAL    .....MPRD 047
C      DIAGONAL    DIAGONAL     DIAGONAL    .....MPRD 048
C .....MPRD 049
C .....MPRD 050
C .....MPRD 051
C SUBROUTINE MPRD(A,B,R,N,M,MSA,MSB,L) .....MPRD 052

```

APPENDIX B

	DOUBLE PRECISION A, B, R	
	DIMENSION A(1),B(1),R(1)	MPRD 053
C		MPRD 054
C	SPECIAL CASE FOR DIAGONAL BY DIAGONAL	MPRD 055
C		MPRD 056
	MS=MSA*10+MSB	MPRD 057
	IF(MS-22) 30,1C,30	MPRD 058
10	DO 20 I=1,N	MPRD 059
20	R(I)=A(I)*B(I)	MPRD 060
	RETURN	MPRD 061
C		MPRD 062
C	ALL OTHER CASES	MPRD 063
C		MPRD 064
30	IR=1	MPRD 065
	DO 90 K=1,L	MPRD 066
	DO 90 J=1,N	MPRD 067
	R(IR)=0	MPRD 068
	DO 80 I=1,M	MPRD 069
	IF(MS) 40,60,4C	MPRD 070
40	CALL LOC(J,I,IA,N,M,MSA)	MPRD 071
	CALL LOC(I,K,IB,M,L,MSB)	MPRD 072
	IF(IA) 50,8C,5C	MPRD 073
50	IF(IB) 70,80,7C	MPRD 074
60	IA=N*(I-1)+J	MPRD 075
	IB=M*(K-1)+I	MPRD 076
70	R(IR)=R(IR)+A(IA)*B(IB)	MPRD 077
80	CONTINUE	MPRD 078
90	IR=IR+1	MPRD 079
	RETURN	MPRD 080
	END	MPRD 081

APPENDIX B

C		LOC 001
C	LOC 002
C		LOC 003
C	SUBROUTINE LOC	LOC 004
C		LOC 005
C	PURPOSE	LOC 006
C	COMPUTE A VECTOR SUBSCRIPT FOR AN ELEMENT IN A MATRIX OF	LOC 007
C	SPECIFIED STORAGE MODE	LOC 008
C		LOC 009
C	USAGE	LOC 010
C	CALL LOC %I,J,IR,N,M,MS<	LOC 011
C		LOC 012
C	DESCRIPTION OF PARAMETERS	LOC 013
C	I - ROW NUMBER OF ELEMENT	LOC 014
C	J - COLUMN NUMBER OF ELEMENT	LOC 015
C	IR - RESULTANT VECTOR SUBSCRIPT	LOC 016
C	N - NUMBER OF ROWS IN MATRIX	LOC 017
C	M - NUMBER OF COLUMNS IN MATRIX	LOC 018
C	MS - ONE DIGIT NUMBER FOR STORAGE MODE OF MATRIX	LOC 019
C	0 - GENERAL	LOC 020
C	1 - SYMMETRIC	LOC 021
C	2 - DIAGONAL	LOC 022
C		LOC 023
C	REMARKS	LOC 024
C	NONE	LOC 025
C		LOC 026
C	SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED	LOC 027
C	NONE	LOC 028
C		LOC 029
C	METHOD	LOC 030
C	MS#0 SUBSCRIPT IS COMPUTED FOR A MATRIX WITH N*M ELEMENTS	LOC 031
C	IN STORAGE %GENERAL MATRIX<	LOC 032
C	MS#1 SUBSCRIPT IS COMPUTED FOR A MATRIX WITH N*%N&1</2 IN	LOC 033
C	STORAGE %UPPER TRIANGLE OF SYMMETRIC MATRIX<. IF	LOC 034
C	ELEMENT IS IN LOWER TRIANGULAR PORTION, SUBSCRIPT IS	LOC 035
C	CORRESPONDING ELEMENT IN UPPER TRIANGLE.	LOC 036
C	MS#2 SUBSCRIPT IS COMPUTED FOR A MATRIX WITH N ELEMENTS	LOC 037
C	IN STORAGE %DIAGONAL ELEMENTS OF DIAGONAL MATRIX<.	LOC 038
C	IF ELEMENT IS NOT ON DIAGONAL %AND THEREFORE NOT IN	LOC 039
C	STORAGE<, IR IS SET TO ZERO.	LOC 040
C		LOC 041
C	LOC 042
C		LOC 043
C	SUBROUTINE LOC(I,J,IR,N,M,MS)	LOC 044
C		LOC 045
	IX=I	LOC 046
	JX=J	LOC 047
	IF(MS-1) 10,20,30	LOC 048
10	IRX=N*(JX-1)+IX	LOC 049
	GO TO 36	LOC 050
20	IF(IX-JX) 22,24,24	LOC 051
22	IRX=IX+(JX*JX-JX)/2	LOC 052
	GO TO 36	LOC 053
24	IRX=JX+(IX*IX-IX)/2	LOC 054
	GO TO 36	LOC 055
30	IRX=0	LOC 056
	IF(IX-JX) 36,32,36	LOC 057
32	IRX=IX	LOC 058
36	IR=IRX	LOC 059
	RETURN	LOC 060
	END	LOC 061

APPENDIX B

C		S INV	10
C	S INV	20
C		S INV	30
C	SUBROUTINE DSINV		
C		S INV	50
C	PURPOSE	S INV	60
C	INVERT A GIVEN SYMMETRIC POSITIVE DEFINITE MATRIX	S INV	70
C		S INV	80
C	USAGE	S INV	90
C	CALL DSINV(A,N,EPS,IER)		
C		S INV	110
C	DESCRIPTION OF PARAMETERS	S INV	120
C	A		
C	- DOUBLE PRECISION UPPER TRIANGULAR PART OF GIVEN		
C	SYMMETRIC POSITIVE DEFINITE N BY N COEFFICIENT		
C	MATRIX.		
C	ON RETURN A CONTAINS THE RESULTANT UPPER	S INV	150
C	TRIANGULAR MATRIX IN DOUBLE PRECISION.		
C	N	S INV	170
C	- THE NUMBER OF ROWS (COLUMNS) IN GIVEN MATRIX.		
C	EPS		
C	- SINGLE PRECISION INPUT CONSTANT WHICH IS USED		
C	AS RELATIVE TOLERANCE FOR TEST ON LOSS OF		
C	SIGNIFICANCE.		
C	IER	S INV	200
C	- RESULTING ERROR PARAMETER CODED AS FOLLOWS		
C	IER=0 - NO ERROR	S INV	210
C	IER=-1 - NO RESULT BECAUSE OF WRONG INPUT PARAME-	S INV	220
C	TER N OR BECAUSE SOME RADICAND IS NON-	S INV	230
C	POSITIVE (MATRIX A IS NOT POSITIVE	S INV	240
C	DEFINITE, POSSIBLY DUE TO LOSS OF SIGNI-	S INV	250
C	FICANCE)	S INV	260
C	IER=K - WARNING WHICH INDICATES LOSS OF SIGNIFI-	S INV	270
C	CANCE. THE RADICAND FORMED AT FACTORIZA-	S INV	280
C	TION STEP K+1 WAS STILL POSITIVE BUT NO	S INV	290
C	LONGER GREATER THAN ABS(EPS*A(K+1,K+1)).	S INV	300
C		S INV	310
C	REMARKS	S INV	320
C	THE UPPER TRIANGULAR PART OF GIVEN MATRIX IS ASSUMED TO BE	S INV	330
C	STORED COLUMNWISE IN N*(N+1)/2 SUCCESSIVE STORAGE LOCATIONS.	S INV	340
C	IN THE SAME STORAGE LOCATIONS THE RESULTING UPPER TRIANGU-	S INV	350
C	LAR MATRIX IS STORED COLUMNWISE TOO.	S INV	360
C	THE PROCEDURE GIVES RESULTS IF N IS GREATER THAN 0 AND ALL	S INV	370
C	CALCULATED RADICANDS ARE POSITIVE.	S INV	380
C		S INV	390
C	SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED	S INV	400
C	DMFSD		
C		S INV	420
C	METHOD	S INV	430
C	SOLUTION IS DONE USING FACTORIZATION BY SUBROUTINE DMFSD.		
C		S INV	450
C	S INV	460
C		S INV	470
C	SUBROUTINE DSINV(A,N,EPS,IER)		
C		S INV	490
C		S INV	500

APPENDIX B

	DIMENSION A(1)	SINV 510
	DOUBLE PRECISION A,DIN,WORK	
C		SINV 530
C	FACTORIZE GIVEN MATRIX BY MEANS OF SUBROUTINE DMFSD	
C	A = TRANSPOSE(T) * T	SINV 550
	CALL DMFSD(A,N,EPS,IER)	
	IF(IER) 9,1,1	SINV 570
C		SINV 580
C	INVERT UPPER TRIANGULAR MATRIX T	SINV 590
C	PREPARE INVERSION-LOOP	SINV 600
	1 IPIV=N*(N+1)/2	SINV 610
	IND=IPIV	SINV 620
C		SINV 630
C	INITIALIZE INVERSION-LOOP	SINV 640
	DO 6 I=1,N	SINV 650
	DIN=1.00/A(IPIV)	
	A(IPIV)=DIN	SINV 670
	MIN=N	SINV 680
	KEND=I-1	SINV 690
	LANF=N-KEND	SINV 700
	IF(KEND) 5,5,2	SINV 710
	2 J=IND	SINV 720
C		SINV 730
C	INITIALIZE ROW-LOOP	SINV 740
	DO 4 K=1,KEND	SINV 750
	WORK=0.00	SINV 760
	MIN=MIN-1	SINV 770
	LHOR=IPIV	SINV 780
	LVER=J	SINV 790
C		SINV 800
C	START INNER LOOP	SINV 810
	DO 3 L=LANF,MIN	SINV 820
	LVER=LVER+1	SINV 830
	LHOR=LHOR+L	SINV 840
	3 WORK=WORK+A(LVER)*A(LHOR)	
C	END OF INNER LOOP	SINV 860
C		SINV 870
	A(J)=-WORK*DIN	SINV 880
	4 J=J-MIN	SINV 890
C	END OF ROW-LOOP	SINV 900
C		SINV 910
	5 IPIV=IPIV-MIN	SINV 920
	6 IND=IND-1	SINV 930
C	END OF INVERSION-LOOP	SINV 940
C		SINV 950
C	CALCULATE INVERSE(A) BY MEANS OF INVERSE(T)	SINV 960
C	INVERSE(A) = INVERSE(T) * TRANSPOSE(INVERSE(T))	SINV 970
C	INITIALIZE MULTIPLICATION-LOOP	SINV 980
	DO 8 I=1,N	SINV 990
	IPIV=IPIV+1	SINV 1000
	J=IPIV	SINV 1010
C		SINV 1020

APPENDIX B

C	INITIALIZE ROW-LOOP	SINV 1030
	DO 8 K=I,N	SINV 1040
	WORK=0.DO	SINV 1050
	LHOR=J	SINV 1060
C		SINV 1070
C	START INNER LOOP	SINV 1080
	DO 7 L=K,N	SINV 1090
	LVER=LHOR+K-I	SINV 1100
	WORK=WORK+A(LHOR)*A(LVER)	
	7 LHOR=LHOR+L	SINV 1120
C	END OF INNER LOOP	SINV 1130
C		SINV 1140
	A(J)=WORK	SINV 1150
	8 J=J+K	SINV 1160
C	END OF ROW- AND MULTIPLICATION-LOOP	SINV 1170
C		SINV 1180
	9 RETURN	SINV 1190
	END	SINV 1200

APPENDIX B

C		MFSD	10
C	MFSD	20
C		MFSD	30
C	SUBROUTINE DMFSD		
C		MFSD	50
C	PURPOSE	MFSD	60
C	FACTOR A GIVEN SYMMETRIC POSITIVE DEFINITE MATRIX	MFSD	70
C		MFSD	80
C	USAGE	MFSD	90
C	CALL DMFSD(A,N,EPS,IER)		
C		MFSD	110
C	DESCRIPTION OF PARAMETERS	MFSD	120
C	A		
C	- DOUBLE PRECISION UPPER TRIANGULAR PART OF GIVEN		
C	SYMMETRIC POSITIVE DEFINITE N BY N COEFFICIENT		
C	MATRIX.		
C	ON RETURN A CONTAINS THE RESULTANT UPPER	MFSD	150
C	TRIANGULAR MATRIX IN DOUBLE PRECISION.		
C	N		
C	- THE NUMBER OF ROWS (COLUMNS) IN GIVEN MATRIX.	MFSD	170
C	EPS		
C	- SINGLE PRECISION INPUT CONSTANT WHICH IS USED		
C	AS RELATIVE TOLERANCE FOR TEST ON LOSS OF		
C	SIGNIFICANCE.		
C	IER		
C	- RESULTING ERROR PARAMETER CODED AS FOLLOWS	MFSD	200
C	IER=0 - NO ERROR	MFSD	210
C	IER=-1 - NO RESULT BECAUSE OF WRONG INPUT PARAME-	MFSD	220
C	TER N OR BECAUSE SOME RADICAND IS NON-	MFSD	230
C	POSITIVE (MATRIX A IS NOT POSITIVE	MFSD	240
C	DEFINITE, POSSIBLY DUE TO LOSS OF SIGNI-	MFSD	250
C	FICANCE)	MFSD	260
C	IER=K - WARNING WHICH INDICATES LOSS OF SIGNIFI-	MFSD	270
C	CANCE. THE RADICAND FORMED AT FACTORIZA-	MFSD	280
C	TION STEP K+1 WAS STILL POSITIVE BUT NO	MFSD	290
C	LONGER GREATER THAN ABS(EPS*A(K+1,K+1)).	MFSD	300
C		MFSD	310
C	REMARKS	MFSD	320
C	THE UPPER TRIANGULAR PART OF GIVEN MATRIX IS ASSUMED TO BE	MFSD	330
C	STORED COLUMNWISE IN N*(N+1)/2 SUCCESSIVE STORAGE LOCATIONS.	MFSD	340
C	IN THE SAME STORAGE LOCATIONS THE RESULTING UPPER TRIANGU-	MFSD	350
C	LAR MATRIX IS STORED COLUMNWISE TOO.	MFSD	360
C	THE PROCEDURE GIVES RESULTS IF N IS GREATER THAN 0 AND ALL	MFSD	370
C	CALCULATED RADICANDS ARE POSITIVE.	MFSD	380
C	THE PRODUCT OF RETURNED DIAGONAL TERMS IS EQUAL TO THE	MFSD	390
C	SQUARE-ROOT OF THE DETERMINANT OF THE GIVEN MATRIX.	MFSD	400
C		MFSD	410
C	SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED	MFSD	420
C	NONE	MFSD	430
C		MFSD	440
C	METHOD	MFSD	450
C	SOLUTION IS DONE USING THE SQUARE-ROOT METHOD OF CHOLESKY.	MFSD	460
C	THE GIVEN MATRIX IS REPRESENTED AS PRODUCT OF TWO TRIANGULAR	MFSD	470
C	MATRICES, WHERE THE LEFT HAND FACTOR IS THE TRANSPOSE OF	MFSD	480
C	THE RETURNED RIGHT HAND FACTOR.	MFSD	490
C	MFSD	510

APPENDIX B

C		MFSD 500
C		MFSD 520
	SUBROUTINE DMFSD(A,N,EPS,IER)	
C		MFSD 540
C		MFSD 550
	DIMENSION A(1)	MFSD 560
	DOUBLE PRECISION DPIV,DSUM,A	
C		MFSD 580
C	TEST ON WRONG INPUT PARAMETER N	MFSD 590
	IF(N-1) 12,1,1	MFSD 600
1	IER=0	MFSD 610
C		MFSD 620
C	INITIALIZE DIAGONAL-LOOP	MFSD 630
	KPIV=0	MFSD 640
	DO 11 K=1,N	MFSD 650
	KPIV=KPIV+K	MFSD 660
	IND=KPIV	MFSD 670
	LEND=K-1	MFSD 680
C		MFSD 690
C	CALCULATE TOLERANCE	MFSD 700
	TOL=ABS(EPS*SVGL(A(KPIV)))	
C		MFSD 720
C	START FACTORIZATION-LOOP OVER K-TH ROW	MFSD 730
	DO 11 I=K,N	MFSD 740
	DSUM=0.DO	MFSD 750
	IF(LEND) 2,4,2	MFSD 760
C		MFSD 770
C	START INNER LOOP	MFSD 780
2	DO 3 L=1,LEND	MFSD 790
	LANF=KPIV-L	MFSD 800
	LIND=IND-L	MFSD 810
3	DSUM=DSUM+A(LANF)*A(LIND)	
C	END OF INNER LOOP	MFSD 830
C		MFSD 840
C	TRANSFORM ELEMENT A(IND)	MFSD 850
4	DSUM=A(IND)-DSUM	
	IF(I-K) 10,5,1C	MFSD 870
C		MFSD 880
C	TEST FOR NEGATIVE PIVOT ELEMENT AND FOR LOSS OF SIGNIFICANCE	MFSD 890
5	IF(SVGL(DSUM)-TOL) 6,6,9	MFSD 900
6	IF(DSUM) 12,12,7	MFSD 910
7	IF(IER) 8,8,9	MFSD 920
8	IER=K-1	MFSD 930
C		MFSD 940
C	COMPUTE PIVOT ELEMENT	MFSD 950
9	DPIV=DSQRT(DSUM)	MFSD 960
	A(KPIV)=DPIV	MFSD 970
	DPIV=1.DO/DPIV	MFSD 980
	GO TO 11	MFSD 990
C		MFSD1000
C	CALCULATE TERMS IN ROW	MFSD1010
10	A(IND)=DSUM*DPIV	MFSD1020
11	IND=IND+1	MFSD1030
C		MFSD1040
C	END OF DIAGONAL-LOOP	MFSD1050
	RETURN	MFSD1060
12	IER=-1	MFSD1070
	RETURN	MFSD1080
	END	MFSD1090

APPENDIX C

OUTPUT FOR THE EXAMPLE

LIST OF VECTORS WITH OUTLIERS IDENTIFIED BY ASTERISKS

J	DELTA	H/R	J	DELTA	H/R
1	0.6901	91.12	34	1.1316	89.42
2	0.0240	96.05	35	0.8436	90.48
3	0.0139	96.32	36	2.7647	85.05
4	0.4924	102.30	37	0.0249	96.03
5	0.3994	92.57	38	0.0203	98.22
6	0.0446	95.64	39	0.0452	95.63
7	0.6192	91.44	40	1.5831	106.36
8	0.3098	93.12	41	0.0170	96.23
9	0.1806	94.08	42	0.7866	90.71
10	0.5239	91.90	43	0.1973	93.94
11	0.0063	96.60	44	0.2601	93.46
12	0.0173	98.14	45	0.2643	93.43
13	0.3266	101.35	46	0.0440	95.65
14	0.1055	99.55	47	0.0087	96.50
15	0.0390	95.74	48	0.4256	101.94
16	0.0244	98.32	49	2.1912	107.98
17	0.1420	99.93	50	0.0968	99.45
18	0.6830	103.21	51	3.7721	111.35
19	0.5996	102.83	52	0.9550	104.31
20	0.1279	99.79	53	0.7815	103.63
21	1.1196	104.90	54	0.3489	101.49
22	0.0368	98.58	55	3.8632	111.52
23	0.6898	103.24	56	7.8031 *	117.56
24	0.9047	104.12	57	3.2634	110.36
25	0.2784	101.03	58	0.0253	98.34
26	0.8715	90.37	59	0.0146	96.30
27	0.7036	103.30	60	4.2446	82.15
28	1.8903	87.15	61	3.9288	82.72
29	2.6336	109.02	62	3.2739	83.98
30	0.0016	96.89	63	0.7131	91.02
31	0.5893	91.58	64	0.1841	94.05
32	2.7238	85.14	65	0.9473	90.08
33	1.9017	87.12	66	0.4621	102.14

* OUTLIER IDENTIFIED AT 0.01 SIGNIFICANCE LEVEL
SAMPLE SIZE IS 66
NC OF OUTLIERS IS 1
DELSTAR = 6.3530

DATA BEFORE IDENTIFICATION OF OUTLIERS			DATA AFTER DELETION OF OUTLIERS	
VARIABLES	MEAN	S.D.	MEAN	S.D.
H/R	97.1806	7.2956	96.8671	6.8897

APPENDIX C

LIST OF VECTORS WITH OUTLIERS IDENTIFIED BY ASTERISKS

J	DELTA	G	J	DELTA	G
1	1.3790	0.70	34	0.8358	0.85
2	1.1829	0.75	35	0.6848	0.90
3	1.0018	0.80	36	0.5488	0.95
4	0.8358	0.85	37	0.4279	1.00
5	0.6848	0.90	38	0.3220	1.05
6	0.5488	0.95	39	0.2311	1.10
7	0.4279	1.00	40	0.1553	1.15
8	0.3220	1.05	41	0.0945	1.20
9	0.2311	1.10	42	0.0487	1.25
10	0.1553	1.15	43	0.0180	1.30
11	0.0945	1.20	44	0.0022	1.35
12	0.0487	1.25	45	0.0016	1.40
13	0.0180	1.30	46	0.0159	1.45
14	0.0022	1.35	47	0.0897	1.55
15	0.0016	1.40	48	0.1491	1.60
16	0.0159	1.45	49	0.2236	1.65
17	0.0453	1.50	50	0.3131	1.70
18	0.0897	1.55	51	0.4177	1.75
19	0.1491	1.60	52	0.5372	1.80
20	0.3131	1.70	53	0.9861	1.95
21	0.5372	1.80	54	1.7951	2.15
22	0.6718	1.85	55	2.0349	2.20
23	0.8215	1.90	56	4.1346	2.55
24	0.9861	1.95	57	4.4947	2.60
25	1.5703	2.10	58	14.1922 *	3.55
26	1.7951	2.15	59	1.0018	0.80
27	2.0349	2.20	60	0.8358	0.85
28	2.3137	0.50	61	0.6848	0.90
29	2.0575	0.55	62	0.5488	0.95
30	1.5901	0.65	63	0.4279	1.00
31	1.3790	0.70	64	0.3220	1.05
32	1.1829	0.75	65	0.0897	1.55
33	1.0018	0.80	66	2.8446	2.35

* OUTLIER IDENTIFIED AT 0.01 SIGNIFICANCE LEVEL
SAMPLE SIZE IS 66
NO OF OUTLIERS IS 1
DELSTAR = 6.3530

DATA BEFORE IDENTIFICATION OF OUTLIERS			DATA AFTER DELETION OF OUTLIERS	
VARIABLES	MEAN	S.D.	MEAN	S.D.
G	1.3773	0.5767	1.3438	0.5128

APPENDIX C

LIST OF VECTORS WITH OUTLIERS IDENTIFIED BY ASTERISKS

J	DELTA	H/R	G	J	DELTA	H/R	G
1	1.4132	91.12	0.70	34	1.2653	89.42	0.85
2	1.5222	96.05	0.75	35	0.9751	90.48	0.90
3	1.3226	96.32	0.80	36	2.8389	85.05	0.95
4	3.1165	102.30	0.85	37	0.5015	96.03	1.00
5	0.7199	92.57	0.90	38	0.6559	98.22	1.05
6	0.6203	95.64	0.95	39	0.2377	95.63	1.10
7	0.6784	91.44	1.00	40	3.4797	106.36	1.15
8	0.4002	93.12	1.05	41	0.0979	96.23	1.20
9	0.2635	94.08	1.10	42	0.9158	90.71	1.25
10	0.5249	91.90	1.15	43	0.2202	93.94	1.30
11	0.1091	96.60	1.20	44	0.3527	93.46	1.35
12	0.1499	98.14	1.25	45	0.4353	93.43	1.40
13	0.6519	101.35	1.30	46	0.1363	95.65	1.45
14	0.1888	99.55	1.35	47	0.1967	96.50	1.55
15	0.0746	95.74	1.40	48	0.4257	101.94	1.60
16	0.0263	98.32	1.45	49	2.4135	107.98	1.65
17	0.1421	99.93	1.50	50	0.3134	99.45	1.70
18	0.7313	103.21	1.55	51	4.1166	111.35	1.75
19	0.6055	102.83	1.60	52	0.9969	104.31	1.80
20	0.3148	99.79	1.70	53	1.1297	103.63	1.95
21	1.1413	104.90	1.80	54	1.8467	101.49	2.15
22	0.7922	98.58	1.85	55	3.9882	111.52	2.20
23	0.9621	103.24	1.90	56	8.0630	117.56	2.55
24	1.1990	104.12	1.95	57	4.9986	110.36	2.60
25	1.6294	101.03	2.10	58	20.3487 *	98.34	3.55
26	6.1920	90.37	2.15	59	1.3188	96.30	0.80
27	2.0352	103.30	2.20	60	4.3616	82.15	0.85
28	2.6309	87.15	0.50	61	4.0832	82.72	0.90
29	11.1150 *	109.02	0.55	62	3.4158	83.98	0.95
30	2.3074	96.89	0.65	63	0.7541	91.02	1.00
31	1.3906	91.58	0.70	64	0.3372	94.05	1.05
32	2.7500	85.14	0.75	65	2.0682	90.08	1.55
33	1.9632	87.12	0.80	66	2.9773	102.14	2.35

* OUTLIER IDENTIFIED AT 0.01 SIGNIFICANCE LEVEL
SAMPLE SIZE IS 66
NO OF OUTLIERS IS 2
DELSTAR = 8.7115

DATA BEFORE IDENTIFICATION OF OUTLIERS			DATA AFTER DELETION OF OUTLIERS	
VARIABLES	MEAN	S.D.	MEAN	S.D.
H/R	97.1806	7.2556	96.9775	7.2544
G	1.3773	0.5767	1.3562	0.5069

REFERENCES

1. Roman, James: Long-Range Program to Develop Medical Monitoring in Flight. The Flight Research Program-I. Aerospace Medicine, vol. 36, no. 6, June 1965, pp. 514-518.
2. Swaroop, Ram; West, Kenneth A.; and Lewis, Charles E., Jr.: A Simple Technique for Automatic Computer Editing of Biodata. NASA TN D-5275, 1969.
3. Anderson, T. W.: An Introduction to Multivariate Statistical Analysis. John Wiley & Sons, Inc., c.1958.
4. Gillies, J. A., ed.: A Textbook of Aviation Physiology. Pergamon Press, c.1965.